Stat 100a, Introduction to Probability.

Outline for the day:

1. Variance of Bernoulli random variables.
2. Poisson random variables.
3. Negative binomial random variables.
4. Binomial and Geometric review problem.
5. Continuous random variables and densities.
6. Uniform random variables.

Exam 2 will be on Wed Feb23, 2pm-3:15pm.

The computer project is due on Sat Mar5, 8:00pm.

Read through chapter 5. ♠ ♣ ♥ ♦

Homework 2 is due Mon Feb14, 2pm. Email to STAT100AW22@stat.ucla.edu.

Leave all answers as decimals, not fractions, for all homeworks.

http://www.stat.ucla.edu/~frederic/100A/W22

Why does Var(X) = pq if X is Bernoulli?

$\mathrm{Var}(X) = E(X^2) - \mu^2.$

$\mu = E(X) = (1)(p) + (0)(q) = p.$

$E(X^2) = (1^2)(p) + (0^2)(q) = p.$

Therefore, $\mathrm{Var}(X) = p - p^2$

$= p(1-p)$

$= pq.$

Poisson random variables, ch 5.5.

Player 1 plays in a very slow game, 4 hands an hour, and she decides to do a big bluff whenever the second hand on her watch, at the start of the deal, is in some predetermined 15 second interval, i.e. with probability ¼.

Now suppose Player 2 plays in a game where about 10 hands are dealt per hour, so he similarly looks at his watch at the beginning of each poker hand, but only does a big bluff if the second hand is in a 6 second interval, i.e. with probability 1/10.

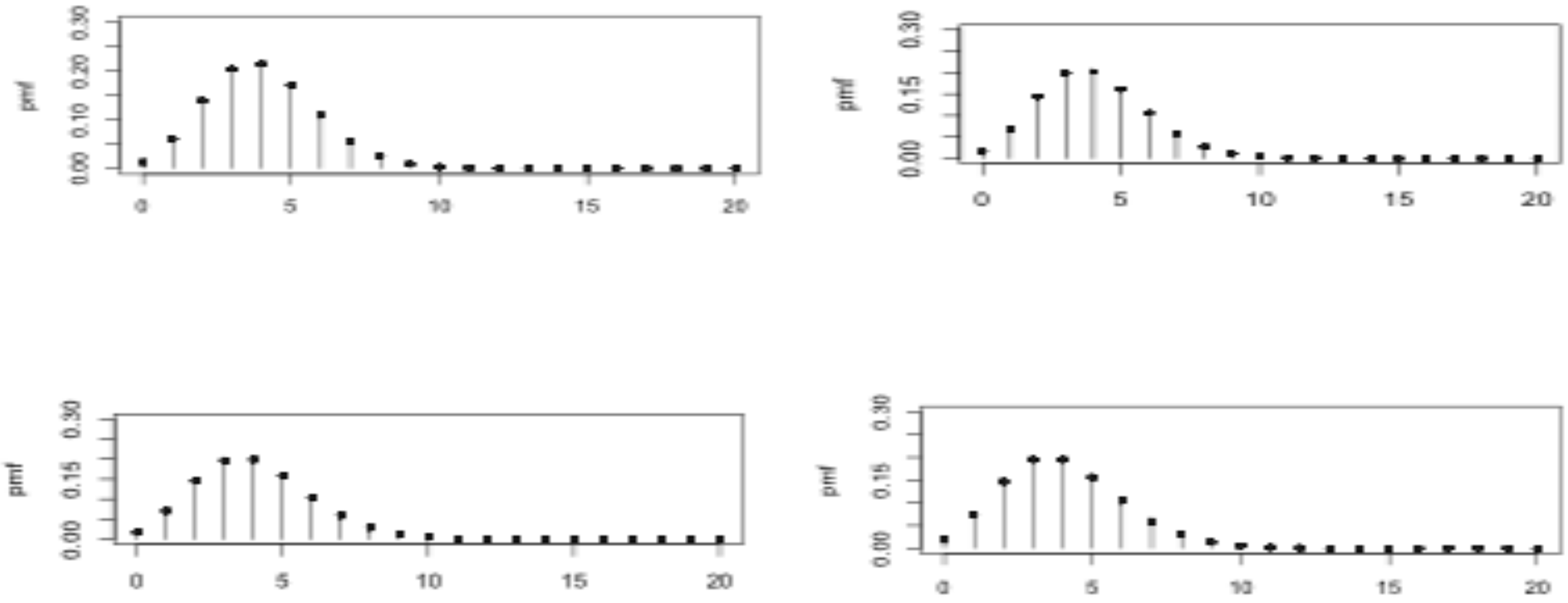
Player 3 plays in a faster game where about 20 hands are dealt per hour, and she bluffs only when the second hand on her watch at the start of the deal is in a 3 second interval, with probability 1/20. Each of the three players will thus average one bluff every hour.

Let X_1, X_2, and X_3 denote the number of big bluffs attempted in a given 4 hour interval by Player 1, Player 2, and Player 3, respectively.

Each of these random variables is binomial with an expected value of 4, and a variance approaching 4.

They are converging toward some limiting distribution, and that limiting distribution is called the *Poisson* distribution.

They are converging toward some limiting distribution, and that limiting distribution is called the *Poisson* distribution. Unlike the binomial distribution which depends on two parameters, n and p, the Poisson distribution depends only on one parameter, λ, which is called the *rate*. In this example, $\lambda = 4$.



The pmf of the Poisson random variable is $f(k) = e^{-\lambda}\lambda^k/k!$, for $k=0,1,2,...$, and for $\lambda > 0$, with the convention that $0!=1$, and where $e = 2.71828....$

The Poisson random variable is the limit in distribution of the binomial distribution as $n \to \infty$ while np is held constant.

For a Poisson(λ) random variable X, $E(X) = \lambda$, and $Var(X) = \lambda$ also. $\lambda = rate$.

Example. Suppose in a certain casino jackpot hands are defined so that they tend to occur about once every 50,000 hands on average. If the casino deals approximately 10,000 hands per day, **a)** what are the expected value and standard deviation of the number of jackpot hands dealt in a 7 day period? **b)** How close are the answers using the binomial distribution and the Poisson approximation? Using the Poisson model, if X represents the number of jackpot hands dealt over this week, what are **c)** $P(X = 5)$ and **d)** $P(X = 5 \mid X > 1)$?

Answer. It is reasonable to assume that the outcomes on different hands are iid, and this applies to jackpot hands as well. In a 7 day period, approximately 70,000 hands are dealt, so X = the number of occurrences of jackpot hands is binomial$(n=70{,}000, p=1/50{,}000)$. Thus **a)** $E(X) = np = 1.4$, and $SD(X) = \sqrt{(npq)} = \sqrt{(70{,}000 \times 1/50{,}000 \times 49{,}999/50{,}000)} \sim 1.183204$. **b)** Using the Poisson approximation, $E(X) = \lambda = np = 1.4$, and $SD(X) = \sqrt{\lambda} \sim 1.183216$. The Poisson model is a very close approximation in this case. Using the Poisson model with rate $\lambda = 1.4$,

c) $P(X=5) = e^{-1.4}\, 1.4^5/5! \sim 1.105\%$.

d) $P(X = 5 \mid X > 1) = P(X = 5 \text{ and } X > 1) \div P(X > 1) = P(X = 5) \div P(X>1) = [e^{-1.4}\ 1.4^5/5!] \div [1 - e^{-1.4}\ 1.4^0/0! - e^{-1.4}\, 1.4^1/1!] \sim 2.71\%$.

Negative binomial random variables, ch5.4.

Recall: if each trial is independent, and each time the probability of an occurrence is p, and X = # of trials until the *first* occurrence, then:

X is Geometric (p), $P(X = k) = p^1 q^{k-1}$, $\mu = 1/p$, $\sigma = (\sqrt{q}) \div p$.

Suppose now X = # of trials until the *rth* occurrence.

Then X = *negative binomial (r,p)*.

e.g. the number of hands you have to play til you've gotten r=3 pocket pairs.

Now X could be 3, 4, 5, …, up to ∞.

pmf: $P(X = k) = \text{choose}(k-1, r-1)\ p^r q^{k-r}$, for k = r, r+1, ….

e.g. say r=3 & k=7: $P(X = 7) = \text{choose}(6,2)\ p^3 q^4$.

Why? Out of the first 6 hands, there must be exactly r-1 = 2 pairs. Then pair on 7th.

P(exactly 2 pairs on first 6 hands) = $\text{choose}(6,2)\ p^2 q^4$. P(pair on 7th) = p.

If X is negative binomial (r,p), then $\mu = r/p$, and $\sigma = [\sqrt{(rq)}] \div p$.

e.g. Suppose X = the number of hands til your 12th pocket pair. P(X = 100)? E(X)? σ?

X = Neg. binomial (12, 5.88%).

$P(X = 100) = \text{choose}(99,11)\ p^{12} q^{88}$

= choose(99,11) * 0.0588 ^ 12 * 0.9412 ^ 88 = **0.104%.**

E(X) = r/p = 12/0.0588 ~ **204**. σ = sqrt(12*0.9412) / 0.0588 = **57.2**.

So, you'd typically *expect* it to take 204 hands til your 12th pair, +/- around 57.2 hands.

Let X = the # of hands until your 1st pair of black aces. What are E(X) and SD(X)?

X is geometric(p), where p = 1/C(52,2) = 1/1326.

E(X) = 1/p = 1326.

SD = (√q) / p, where q = 1325/1326. SD = 1325.5.

What is P(X = 12)?

$q^{11}p$ = 0.0748%.

You play 100 hands. Let X = the # of hands where you have 2 black aces. What is E(X)? What is P(X = 4)?

X is binomial(100,p), where p = 1/1326.

E(X) = np = .0754.

P(X = 4) = C(100,4) p^4 q^{96} = .000118%.

Continuous random variables and their densities, ch6.1.

Density (or pdf = Probability Density Function) f(y):

$$\int_B f(y)\, dy = P(X \text{ in } B).$$

If F(c) is the cumulative distribution function, i.e. $F(c) = P(X \le c)$,

then $f(c) = F'(c)$.

The survivor function is $S(c) = P(X > c) = 1 - F(c)$.

Expected value, $\mu = E(X) = \int y\, f(y)\, dy$. ($= \sum y\, P(y)$ for discrete X.)

For any function g, $E(g(X)) = \int g(y)\, f(y)\, dy$. For instance $E(X^2) = \int y^2 f(y) dy$.

Variance, $\sigma^2 = V(X) = Var(X) = E(X-\mu)^2 = E(X^2) - \mu^2$.

$SD(X) = \sqrt{V(X)}$.

For examples of pictures of pdfs, see p104, 106, and 107.

Uniform example.

Recall for a continuous random variable X,
the pdf f(y) is a function where $\int_a^b f(y)dy = P\{X \text{ is in } (a,b)\}$,
$E(X) = \mu = \int_{-\infty}^{\infty} y\, f(y)dy$,
and $\sigma^2 = Var(X) = E(X^2) - \mu^2$. sd(X) = σ.
If X is a continuous rv, then $P(X \le a) = P(X < a)$, because $P(X = a) = \int_a^a f(y)dy = 0$.
If X is uniform(a,b), then f(y) = 1/(b-a) for y in (a,b), and f(y) = 0 otherwise.

For example, if X is uniform (100,120),

then f(y) = 1/20 for y in (100,120), and f(y) = 0 otherwise.

$E(X) = \int_{-\infty}^{\infty} y\, f(y)dy = \int_{100}^{120} y\, (1/20)\, dy = 1/20\, (120^2/2 - 100^2/2) = 110.$

$E(X^2) = \int_{-\infty}^{\infty} y^2\, (1/20)dy = 1/20\, (120^3/3 - 100^3/3) = 12133.33.$

$Var(X) = E(X^2) - \mu^2 = 12133.33 - 110^2 = 33.33.$

SD(X) = 5.77.

Min of Uniforms example.

Suppose X and Y are independent uniform random variables on (0,1), and Z = *min*(X,Y). **a)** Find the pdf of Z. **b)** Find E(Z). **c)** Find SD(Z).

a. For c in (0,1), $P(Z > c) = P(X > c \;\&\; Y > c) = P(X > c)\,P(Y > c) = (1-c)^2 = 1 - 2c + c^2$.
So, $P(Z \le c) = 1 - (1 - 2c + c^2) = 2c - c^2$.
Thus, $\int_0^c f(c)dc = 2c - c^2$. So f(c) = the derivative of $2c - c^2 = 2 - 2c$, for c in (0,1).
Obviously, f(c) = 0 for all other c.
b. $E(Z) = \int_{-\infty}^{\infty} y\, f(y)dy = \int_0^1 c\,(2-2c)\, dc = \int_0^1 2c - 2c^2\, dc = c^2 - 2c^3/3]_{c=0}^{1}$
$= 1 - 2/3 - (0 - 0) = 1/3$.
c. $E(Z^2) = \int_{-\infty}^{\infty} y^2\, f(y)dy = \int_0^1 c^2\,(2-2c)\, dc = \int_0^1 2c^2 - 2c^3\, dc = 2c^3/3 - 2c^4/4]_{c=0}^{1}$
$= 2/3 - 1/2 - (0 - 0) = 1/6$.
So, $\sigma^2 = Var(Z) = E(Z^2) - [E(Z)]^2 = 1/6 - (1/3)^2 = 1/18$.
$SD(Z) = \sigma = \sqrt{(1/18)} \sim 0.2357$.